The Objective Structured Clinical Examination and student collusion

TEACHING AND LEARNING ETHICS

The Objective Structured Clinical Examination and student collusion: marks do not tell the whole truth

R Parks, P M Warren, K M Boyd, H Cameron, A Cumming, G Lloyd-Jones

J Med Ethics 2006;**32**:734–738. doi: 10.1136/jme.2005.015446

See end of article for authors' affiliations

Correspondence to: G Lloyd-Jones, Centre for Postgraduate Learning and Teaching, Building 6, Cranfield University, Cranfield, Bedfordshire MK43 0AH, UK; G.Lloyd-Jones@cranfield.ac.uk

Received 2 December 2005
Revised 6 January 2006
Accepted 9 January 2006

Objective: To determine whether the marks in the third year Objective Structured Clinical Examination (OSCE) were affected by the collusion reported by the students themselves on an electronic discussion board.
Design: A review of the student discussion, examiners' feedback and a comparison of the marks obtained on the 2 days of the OSCE.
Participants: 255 third year medical students.
Setting: An OSCE consisting of 15 stations, administered on three sites over 2 days at a UK medical school.
Results: 40 students contributed to the discussion on the electronic discussion board. The main points raised were perceived inequity between students who did, or did not, have prior knowledge of the station content, and the lack of honesty and professionalism of their peers. Most contributors claimed to have received, or knew of others receiving, prior knowledge, but none confessed to passing on information. No significant difference ($p = 0.16$) was observed in the overall mark for the OSCE on day 1 (mean 390 (SD 37)) and day 2 (mean 397 (38)). On day 2, marks were considerably greater for four stations and markedly lower for three stations. It was not obvious why collusion should affect these station marks. A clear indication of the effects of collusion could only be obtained from a single subsection of an individual station (pathology) where 82 students on day 2 incorrectly gave the diagnosis from day 1.
Conclusion: Marks do not provide a sound inference of student collusion in an OSCE and may mask the aspects of professional development of students.

Higher education institutions are paying increased attention to academic dishonesty among their students. Although cheating under any circumstances is a serious regulatory matter, its importance on courses leading to professional status is arguably greater. The General Medical Council states that doctors should be honest and trustworthy; so a question mark hangs over the suitability of any student found guilty of cheating for entry into the profession.[1] Because little is known about the career trajectories of students who transgress ethical codes, a link between student dishonesty and future professional lapses cannot be assumed, however plausible the association may seem. Limited evidence from self-reported data from engineering students in the US points to an association between cheating in high school or college and in the workplace.[2] Universities therefore experience dilemmas when dealing with instances of cheating in medical schools.

For obvious reasons, it is difficult to estimate the frequency of occurrence of clandestine behaviour such as cheating.[3] On a limited evidence base, cheating among medical students appears to be neither new nor confined to particular cultures,[3–10] although the introduction of information technology, cultural norms[3 7] and mismatches between student and faculty perceptions[11] have been claimed as contributory. In the main, reports either describe specific incidents[9 10] or depend on self-reported data from questionnaire surveys that reflect student experience and intention.[5–8] In some studies, the frequency of observations of cheating, or intentions to cheat, increased with seniority, suggesting that increasing experience of medical school is accompanied by the development of more relaxed peer norms towards academic misconduct.[6–8]

In 2000, the *BMJ* reported the management of a student caught cheating during a clinical finals examination at a London medical school.[10] Among the many responses was an anonymous confessional, from ''recent graduates'' of a UK medical school, describing collusion on a major scale in relation to an Objective Structured Clinical Examination (OSCE).[12] Students had gained sight of a marking sheet 2 days before the exam and disseminated the information, enabling recipients to revise selectively for the OSCE. A greater number of merits and distinctions were awarded that year.

There is genuine uncertainty about whether advance notice of content advantages a candidate sitting a performance assessment. The evidence from the literature is far from clear. Several studies have examined scores when similar OSCEs are administered over periods varying from days to 3 months and have shown that overall score trends do not rise over time.[13–18] In two of these studies, accompanying self-reported data confirmed that students had discussed station content between administrations of OSCEs, although the nature of the interactions was not investigated.[13 17] In another study, in which students were officially encouraged to discuss exam content, there was no evidence of rising scores over time.[19] Contrary reports exist, however, of rising trends in overall marks with successive OSCE administrations[20 21] and, in a repeat of their initial study,[14] Colliver *et al*[22] analysed within station scores and showed that some component items did exhibit rising trends even when overall scores did not. In cases where OSCE stations have been reused over longer intervals of a year, the evidence supports rising scores over time.[23 24]

The assumptions underlying these studies require closer scrutiny. Some assume that advance notice of station content will advantage candidates, resulting in higher scores for them

Abbreviation: OSCE, Objective Structured Clinical Examination

than their peers. From this arises the claim that a trend for a rising score in a cohort reflects a potential breach of exam security and, conversely, that finding consistent scores over time indicates that no major violation of exam security has occurred.[13 14 16–19] The alternative view, that advance information can disadvantage candidates, resulting in lower marks over time, is considered, but features less prominently.[15 20 22] If information about content has the potential to either advantage or disadvantage a candidate, then inferences relating to trends in overall marks of the peer group to collusion become unreliable. Even consistent mean marks over time cannot be taken to indicate that exam security has been maintained.[22] Detecting breaches of test security, particularly at the individual level, becomes difficult, if not impossible, under these circumstances. In addition, a focus on marks has seemingly diverted attention from the ethical implications of collusion and its potential significance for the development of professionalism in undergraduate students.

AIMS

Against this background, we describe a critical incident that occurred in a UK medical school, in which student collusion at a third year OSCE was seen on a staff–student electronic discussion board. We consider the incident to be of wider relevance and interest owing to the many issues that it raises for students, teachers and educational institutions, in which the OSCE format is used in the assessment programme.

We aim to:

- describe the incident, its management and consequences for the school;
- challenge the view that trends in overall means or station marks, derived from sequential OSCE sittings, provide a sound basis for the presence or absence of collusion; and
- discuss ethical issues arising out of collusion from student and staff perspectives.

Because this paper relates to events in real time, which do not conform easily to the predetermined character of a research study, the following structure has been chosen:

- a description of the incident in context;
- a description of the staff response to the incident; and
- an analysis of OSCE marks.

CRITICAL INCIDENT

OSCE implementation

In all, 255 third year students sat for a summative OSCE consisting of 15 stations of 10 min each. The exam was conducted at three different sites over 2 consecutive days, resulting in three sittings per day at each site. Table 1 describes the individual station tasks and any intended variations between days 1 and 2. Marking sheets were standardised and the standard for scores was set using the Angoff method.

Students taking the OSCE on day 1 were explicitly advised not to pass on information about the exam to students at later sittings. Steps were taken to avoid successive groups of candidates talking to one another, although it was impossible to prevent day 1 and 2 students communicating during the assessment period.

Electronic discussion board

The medical school runs a series of password-protected, electronic discussion boards, organised by year of study, which are accessible to staff and students in the school. Less than 2 h after the end of the day 2 OSCE, students initiated a discussion about collusion between day 1 and 2 students, which continued until the publication of results 12 days later. Nearly 40 students contributed, about three quarters of whom remained anonymous.

Table 1 Description of the Objective Structured Clinical Examination stations in order of question consistency between days 1 and 2

Station blueprint	Station task	Question consistency on days 1 and 2
Gastrointestinal system	Physical examination	Identical
Locomotor system	Physical examination	Identical
Cardiovascular system	Physical examination	Identical
Respiratory system	History taking	Identical
Locomotor system: diagnosis and management	Video Written Q and A	Identical
Pharmacology	Prescribing for different scenarios on medicines sheet	Identical
Ethics	Discuss smoking cessation during pregnancy	Identical
Practical skills	Connecting intravenous fluids	Identical
Communication skills	Diagnosis from written scenario and interview about alcohol consumption	Different scenario and diagnosis on days 1 and 2
Resuscitation	Basic life support	Half was identical in addition Day 1: defibrillation Day 2: airway management
Pathology	Written Q and A on specimens	Identical questions: specimens varied
Anatomy	Written Q and A on scenarios and specimens and *x* rays	Identical questions: specimens varied
Microbiology	Written Q and A on cultures, *x* rays and treatment	Identical questions: specimens varied
ECG	Written Q and A on interpretation of electrocardiograms	Different questions and traces
Radiology	Interpretation of *x* rays	Different questions and *x* rays

ECG, electrocardiogram; Q and A, questions and answers.

The main topic was the perceived lack of equity between those students who had, and had not, received insider information about OSCE content. Students without insider information argued that colluding students retained an advantage over the rest of their colleagues, which devalued the achievements of honest students and the assessment process. There was concern that honesty was not being rewarded. Several posts criticised their peer's dishonesty and its potential effect on professionalism and future practice. Although students reported either receiving insider information themselves or knowing of someone who had, no one confessed to conveying information. Those receiving insider information disclaimed any advantage, but others reported studying relevant electronic learning resources and memorising names of drug in the interim. Clearly, collusion had occurred between students sitting the OSCE on the same day, as well as on different days.

Staff response

Under the circumstances, staff thought it imperative to respond to events, as the electronic discussion board is publicly accessible to the entire school. Although students who use the facility may elect to remain anonymous, staff reserves the right to uncover the identity of any student who fails to maintain proper standards. Staff associated with the

OSCE vigorously debated appropriate responses to the incident. Reprimanding students who had admitted receiving advance information was considered, but was narrowly rejected on the following grounds. Firstly, no student had admitted passing on information and these students were arguably as guilty as those who received information. Also, students would probably not divulge the names of informants. Secondly, the discussion board thread probably represented only the tip of the iceberg and the collusion may have been more widespread. Further problems arose from the interpretation of electronic interaction: without clues of non-verbal behaviour, irony and sarcasm are difficult to differentiate from factual statements.

Staff recognised that targeting the students who contributed to the discussion would make scapegoats of a few, whereas others, perhaps many more, would evade censure. Instead, they chose to post a response to all third year students on the electronic discussion board, dealing with some of the students' misconceptions about the implementation of the OSCE, recognising the inequity of the situation and promising to work towards a form of implementation that reduced opportunities for collusion to uphold the honest students. It was emphasised that the policy of non-disclosure of exam content would continue to remain an offence under the current assessment regulations.

RESULTS ANALYSIS

In the light of the student discussion, we reviewed the OSCE marks. The mean and SD for each day, differences in means (day 2−day 1) and the 95% confidence interval for the difference were calculated for the overall mark and the marks for each station (table 2). The day 1 and 2 marks were compared using the two sample t test (SPSS V.13). Although multiple comparisons were carried out, a Bonferroni correction was considered to be inappropriate because of the lack of independence of the overall mark and the station marks. Caution is therefore required when interpreting the reported p values.

The small increase in the mean overall mark between days 1 and 2 was not significant at the 0.05 level. For individual stations, roughly half showed significant differences between the average marks on the two days. It was impossible to attribute these findings to collusion, however, because other influences contributed to the average station mark. These included variations in examiners, student ability and case specificity arising from the participation of different patients at different sittings.[25] Secondly, analysis of the results from all students sitting the exam on a specific day reflects an average effect for the class. If only a few students colluded then the effect would be masked by the results from the large majority who did not. Thirdly, collusion can affect the mark positively or negatively and the overall effect is diluted when considering the average values. Finally, the mark for an individual station was the sum of multiple sections, each of which may have been affected in opposite ways by collusion. For instance, a student may only lose or gain 2 marks out of 40 for a specific item. Thus, analysis of the marks for individual stations was not reliable for detecting or quantifying collusion.

We therefore decided to look in detail at the marks from those stations where examiners had commented spontaneously, without prompting by staff inquiries, that they had suspicions of collusion for their particular station (Communication skills, Pathology).

Communication skills

Examiners of the Communication skills station commented verbally to one of the authors (GL-J) that they had noted that day 2 students were offering the day 1 diagnosis, despite changes to the station scenario. Attempts to analyse the component items of the Communication skills answer sheets were foiled because the marks awarded for supplying the correct diagnosis were aggregated with marks for other aspects of the student-simulated patient interaction. It was therefore not possible to either support or refute the examiners' assertions of collusion.

Pathology

We were alerted to possible collusion in the Pathology station by the examiners' formal feedback comments, which were posted on the third year website. The examiners stated that ''It seems possible that collusion was a confounding factor in day 2 answers to the second part of the question. The histology picture changed from tuberculosis on day 1 to squamous carcinoma on day 2.''

At this station, the scenario, questions and marking schedule were unchanged on days 1 and 2; only the photomicrograph specimen was changed (day 1 tuberculosis

Table 2 Total and stations scores for days 1 and 2

Station	Day 1	Day 2	Mean differences (95% CI)	p Value
n	128	127		
Total score	389.9 (36.8)	396.5 (38.3)	6.6 (−2.66 to 15.9)	0.162
Stations with questions consistent between days				
Gastrointestinal system	30.5 (4.4)	31.1 (4.7)	0.7 (−0.4 to 1.8)	0.236
Locomotor system	28.8 (6.8)	28.0 (6.6)	−0.9 (−2.5 to 0.8)	0.311
Cardiovascular system	30.4 (4.7)	28.5 (5.0)	−1.8 (−3.0 to −0.6)	0.003
Respiratory system	27.8 (4.6)	29.4 (4.8)	1.6 (0.5 to 2.8)	0.006
Locomotor system: diagnosis and management	25.6 (5.8)	29.7 (5.4)	4.1 (2.7 to 5.5)	0.000
Pharmacology	16.1 (6.0)	17.0 (6.8)	0.8 (−0.8 to 2.4)	0.314
Ethics	27.8 (4.7)	27.1 (5.0)	−0.7 (−1.9 to 0.4)	0.220
Practical skill	27.3 (6.1)	27.7 (6.4)	0.4 (−1.2 to 1.9)	0.643
Stations with questions partially changed between days				
Communication skills	29.1 (5.4)	26.7 (5.5)	−2.4 (−3.8 to −1.01)	0.001
Resuscitation	30.0 (5.3)	32.0 (4.4)	2.0 (0.7 to 3.2)	0.002
Stations with questions changed between days				
Pathology	26.4 (6.4)	24.5 (5.3)	−1.9 (−3.4 to −0.5)	0.010
Anatomy	19.7 (4.5)	24.0 (4.4)	4.3 (3.2 to 5.4)	0.000
Microbiology	19.6 (6.9)	20.8 (6.5)	1.1 (−0.5 to 2.8)	0.179
ECG	26.4 (7.5)	26.2 (6.0)	−0.2 (−1.9 to 1.5)	0.820
Radiology	24.1 (5.1)	23.6 (5.3)	−0.5 (−1.7 to 0.8)	0.483

Data are given as means and (SD) unless otherwise specified.
ECG, electrocardiogram.

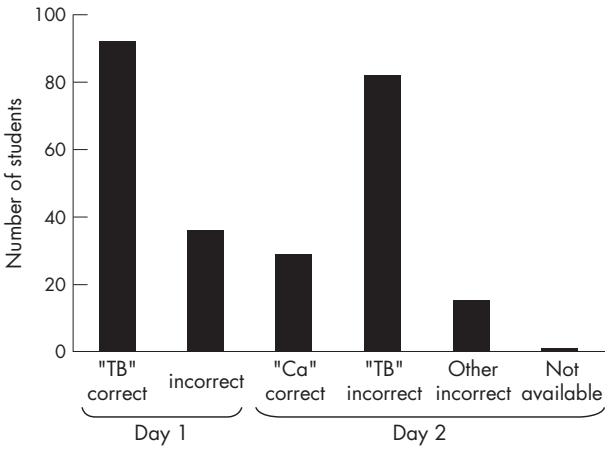

Figure 1 Number of correct or incorrect answers for diagnosis given at the Pathology station on days 1 and 2. Day 1 ''TB'' correct, number giving the correct answer of tuberculosis; day 1 Incorrect, number giving any incorrect answer; day 2 ''Ca'' correct, number giving the correct answer of squamous carcinoma; day 2 ''TB'' incorrect, number giving tuberculosis as an incorrect answer; day 2 Other incorrect, number giving an incorrect answer other than tuberculosis; not available, script on original mark sheet indecipherable.

and day 2 squamous carcinoma). The number of correct and incorrect answers for the diagnosis given each day was counted, the results of which are summarised in fig 1. The findings are congruent with collusion between day 1 and day 2 students.

DISCUSSION

These findings challenge the view that trends in overall mean or station marks, derived from sequential OSCE sittings, provide a sound assumption for the presence or absence of collusion. Instead, they support the later work of Colliver *et al*[22] in showing the need to examine specific station component items, when searching for breaches of test security. The incident also supports the contention that collusion has the power to disadvantage rather than to benefit students who indulge in collusion.[15 20] The two premises, that overall marks are insufficiently sensitive as indicators of collusion, and the uncertain effect of advance notice on marks, go some way in explaining the variations in previous studies.

Some may criticise our conclusions on the grounds that the Pathology question preferentially tested knowledge rather than skills, as knowledge is easier to reproduce at short notice. There is some truth in this, but it is only partial. Although the OSCE is promoted as a skill-based assessment, skill must be constructed on, and integrated with, knowledge if it is to be of value in educating the safe, competent professional. Representing skills and knowledge as entirely different entities is unhelpful when, in reality, the relationship is far more complex. Furthermore, the anecdotal evidence from the Communication skills examiners shows how a station, perceived primarily as a skills test, is not immune to collusion.

It is reassuring that the marks awarded for the specific items are small and, standing alone, would have little effect on the outcome. It is, however, possible for multiple small breaches to have a cumulative effect for an individual student, which would probably go undetected. From the perspectives of the staff and honest students, there is comfort in knowing that malpractice was apparently not rewarded in this case. Yet, the incident threw up difficult issues for staff in handling the public, yet anonymous, electronic report and in protecting honest students. This is probably not an isolated case. When this incident was reported at a recent international conference in medical education, many members of the audience from around the world related similar experiences.

We describe below amendments to the OSCE aimed at reducing opportunities for collusion but, in reality, it will be impossible to block the flow of information, particularly electronic texting and emails. Trusting students will always be necessary when OSCEs are repeated day after day. Although plagiarism can be detected by software, detection of collusion in performance examinations presents an altogether more difficult problem. The deliberate inclusion of questions designed to test the possibility of collusion is one option for consideration. Colliver *et al*[22] follow a similar strategy in suggesting an examination of working diagnosis scores in the context of a simulated-patient examination in the US.

Steps have since been taken to modify the OSCE format to reduce opportunities for collusion. The OSCE is now divided into two halves, each one being administered on one day only. One half consists of clinically based stations, involving interaction with simulated and real patients; the other is delivered via computer and includes questions on patient management, data handling and analysis, and interpretation of investigations. Although this will not exclude the possibility of collusion, it will reduce the opportunities for it.

For institutions, the more difficult problem is upholding honesty and protecting students who do not cheat, for they can feel disregarded by the system. From this incident and other reports, it seems that students are concerned about collusion and cheating, but may lack opportunities to voice their opinions.[10 12] Possibly, the anonymous protection afforded by the discussion board may have encouraged students to speak out in this instance. To discourage cheating, Glick[3] proposes that schools carefully articulate their expectations of students and take positive, public steps to encourage a culture in which unethical behaviour simply becomes unacceptable. Otherwise, there is the danger that turning a blind eye, however unintentional, may encourage the growth of undesirable norms in which peer group loyalty overtakes professional ethics. It is in this spirit that we report the incident, to publicise the uncertain effects of collusion in relation to OSCEs and to open a debate about these issues in relation to current OSCE implementation practices.

CONCLUSION

Although research examining the effects of collusion on OSCE marks has been conducted, there is little information about the nature of collusion and its ethical implications. An incident of collusion at a UK school showed that it could disadvantage students, although its effect on marks appeared to be limited. Honest students raised ethical concerns about cheating colleagues and the possible effect on their future professional practice. In the absence of any evidence against individual students, the school issued a public statement and has since modified the implementation of the OSCE to reduce opportunities for cheating. The potential influence of these issues on the development of social and cultural norms in medical school culture has been considered. In view of the global popularity of the OSCE format for assessing professional competence, these findings may have wider relevance.

ACKNOWLEDGEMENTS

We thank Mr Robert Lee of the Medical Statistics Unit, University of Edinburgh, UK, for advice on the statistics applied to the data given in table 2.

Authors' affiliations

R Parks, Chancellors Building, Little France, Edinburgh, UK
P M Warren, H Cameron, Medical Teaching Organisation, University of Edinburgh, Edinburgh
K M Boyd, Medical School Buildings, University of Edinburgh
A Cumming, College of Medicine and Veterinary Medicine, University of Edinburgh
G Lloyd-Jones, Cranfield Professional Development, Cranfield University, Bedfordshire, UK

Competing interests: None.

REFERENCES

1 **General Medical Council**. Duties of a doctor. http://www.gmc-uk.org/guidance/library/duties of a doctor.asp (accessed 19 October 2006).
2 **Harding TS**, Carpenter DD, Finelli CJ, *et al.* Does academic dishonesty relate to unethical behaviour in professional practice? An exploratory study. *Sci Eng Ethics* 2004;**10**:311–24.
3 **Glick S**. Cheating at medical school. *BMJ* 2001;**322**:250–1.
4 **Becker HS**, Geer B, Hughes EC, *et al. Boys in white*. Chicago: Chicago University Press, 1977.
5 **Simpson DE**, Yindra KJ, Towne JB, *et al.* Medical students' perceptions of cheating. *Acad Med* 1989;**64**:221–2.
6 **Rennie SC**, Crosby JR. Are ''tomorrows doctors'' honest? Questionnaire study exploring medical students' attitudes and reported behaviour on academic misconduct. *BMJ* 2001;**322**:274–5.
7 **Hrabak M**, Vujaklija A, Vodopivec I, *et al.* Academic misconduct among medical students in a post-communist country. *Med Educ* 2004;**38**:276–85.
8 **Trivino X**, Sirhan M, Wright A, *et al.* Cheating on tests among medical students. *Poster Presentation at AMEE Conference*. Edinburgh, 2004.
9 **Osborn E**. Punishment: a story for medical educators. *Acad Med* 2000;**75**:241–4.
10 **Smith R**. Cheating at medical school. *BMJ* 2000;**321**:398.
11 **Ginsberg S**, Regehr G, Stern D, *et al.* The anatomy of the professional lapse: bridging the gap between traditional frameworks and students' perceptions. *Acad Med* 2002;**77**:516–22.
12 **Anonymous**. Rapid responses to cheating at medical school. *BMJ* 7/9/2000. http://bmj.bmjjournals.com/cgi/eletters/321/7258/398 (accessed 14 January 2005).
13 **Williams RG**, Lloyd D, Simonton DK. Sources of OSCE examination and perceived helpfulness: a study of the grapevine. In: Hart IR, Harden RM, Des Marchais J, eds. *Current developments in assessing clinical competence*. Montreal: Can-Heal Publications, 1990.
14 **Colliver J**, Barrows H, Nu Viet Vu C, *et al.* Test security in examinations that use standardized-patient cases at one medical school. *Acad Med* 1991;**66**:279–82.
15 **Stillman PL**, Haley HA, Sutnick AJ, *et al.* Is test security an issue in a multistation clinical assessment? A preliminary study. *Acad Med* 1991;**66**:S25–7.
16 **Rutala PJ**, Witzke DB, Leko EO, *et al.* Sharing of information by students in an objective structured clinical examination. *Arch Int Med* 1991;**151**:541–4.
17 **Skakun EN**, Cook DA, Morrison JC. Test security on sequential OSCE and multiple choice examinations. In: Harden R, Hart I, Mulholland H, eds. *Approaches to the assessment of clinical competence*. Norwich: Page Brothers, 1992.
18 **Niehaus AH**, DaRosa DA, Markwell SJ, *et al.* Is test security a concern when OSCE stations are repeated across clerkship rotations? *Acad Med* 1996;**71**:287–9.
19 **Swartz M**, Colliver J, Cohen D, *et al.* The effect of deliberate, excessive violations of test security on a standardized-patient examination: an extended analysis. *Sixth Ottawa Conference on Medical Education*. Toronto: University of Toronto Bookstore Custom Printing, 1995:280–4.
20 **Wilkinson TJ**, Newble DI, Don Wilson P, *et al.* Development of a three centre simultaneous objective structured clinical examination. *Med Educ* 2000;**34**:798–807.
21 **Mavis BE**. Does studying for an objective structured clinical examination make a difference? *Med Educ* 2000;**34**:808–12.
22 **Colliver JA**, Travis TA, Robbs RS, *et al.* Test security in standardized-patient examinations: analysis with scores on working diagnosis and final diagnosis. *Acad Med* 1992;**67**:S7–9.
23 **Cohen R**, Rothman AI, Ross J, *et al.* Impact of repeated use of objective structured clinical examination stations. *Acad Med* 1993;**68**:S73–5.
24 **Jolly B**, Newble D, Chiner T. Learning effects of reusing stations in an objective structured clinical examination. *Teach Learn Med* 1993;**5**:66–71.
25 **Newble D**, Dauphinee D, Macdonald M, *et al.* Guidelines for assessing clinical competence. *Teach Learn Med* 1994;**6**:213–20.